STATISTICAL LITERACY AND EVIDENTIAL STATISTICS

Milo Schield, Augsburg College
Dept. of Business & MIS. 2211 Riverside Drive. Minneapolis, MN 55454

Abstract: Statistical literacy is becoming a graduation requirement for many reasons. Yet most introductory courses focus just on descriptive statistics and statistical inference. In so doing, the importance of probability and chance (statistical inference) is over-emphasized while the role of bias and the importance of interpreting tables and charts is under-emphasized. To be statistically literate, students must also understand probability as measuring strength of belief (Bayesian) and the use of observational studies to strengthen belief in causality (evidential statistics). Evidential statistics studies the use of statistics as evidence in supporting the truth of claims involving causal explanations and predictions. Studying evidential statistics is helpful in understanding statistical arguments in business, medicine, insurance, public affairs, law and epidemiology. Evidential statistics deals with the ambiguities of language and subtle distinctions involving conditional logic and the role of chance. Functionally, the relation between traditional statistics and evidential statistics is like that between micro and macroeconomics or between formal logic and practical reasoning (critical thinking). Students need an introduction to both traditional and evidential statistics. Sources of evidential statistics are reviewed along with arguments for and against teaching this new subject.

Keywords: Teaching, Epistemology, Philosophy of Science, Strength of Belief, Observational Studies

1. Reasons for Statistical Literacy

Statistical literacy is becoming a graduation requirement for college students. Statistical literacy is being required for several reasons:

Statistics are commonly used in arguments about social policy. Issues such as poverty, welfare, distribution of wealth, crime and drugs are all presented using statistics. Students who are statistically literate should be better able to read and interpret the statistics used in broad social arguments such as *The Bell Curve* (Herrnstein and Murray) or *The Ultimate Resource* (Simon).

Students have difficulty evaluating arguments involving statistics. Statistics are often viewed as numbers just like arithmetic where mathematical operations are always true: 2 + 2 = 4. So if cars with car phones have more accidents than cars without, students conclude that if one installs a car phone one can expect an increased risk of having an accident.

Students have difficulty giving alternate explanations for observed statistical associations. If direct causation is a plausible explanation for an observed association, students often mistakenly conclude that this correlation is a sign of direct causation. Suppose that cars with mobile phones have a higher rate of accidents than cars without. Students have difficulty imagining a confounding factor such as age: younger drivers are more likely to have mobile phones and younger drivers are more likely to have accidents.

Students have difficulty evaluating the strength of an argument involving statistics as premises. In reading *The Bell Curve*, graduate students typically focused on the reliability of the data rather than on the amount of support provided by the summaries of the data (assuming those summaries were true). Students often presumed that if the summary is true, then the causal inference being supported must be true.

2. Contemporary Statistics

Today, most introductory courses involve two parts: descriptive statistics and inferential statistics.

Descriptive statistics deals with summarizing and modeling data–with reducing variability physically (better studies) or conceptually (better models).

Inferential statistics studies chance in order to answer, "How likely is an outcome if due entirely to chance?" If this association is unlikely if due to chance, then we say, "This observation is 'statistically significant'."

As for the relation between association and causation, the common assertion is a simple negative statement, "association is not causation." Most texts and teachers have a number of examples to illustrate this point.

This contemporary approach to introductory applied statistics is inadequate to provide students with on-going statistical literacy. Emphasizing data is necessary–but not sufficient–to teach statistical literacy.

3. Problems in Descriptive Statistics

In teaching statistical literacy, descriptive statistics does not give adequate attention to the understanding and communication of basic ideas.

Many students have problems describing percents. They mistakenly presume if "20% of females are runners", then "20% of runners are females" or "The percentage of females among runners is 20%." Some mistakenly presume, "the percentage of runners who smoke is twice as great among males as females" is different than "the percentage of smokers is twice as great among male runners as female runners."

Many students have difficulty describing simple numeric comparisons. They mistakenly presume that "20 is two times bigger than 10" or "if A is 200% bigger than B", then A is twice as big as A." They mistakenly presume that if interest rates go from 5% to 10%, they have increased by 5%.

Although students may be warned about bias and unrepresentative samples, there is little attention given to the power of confounding factors to change the magnitude (and even the direction) of an observed association between two variables. Simpson's Paradox is either omitted or else relegated to an optional section or to a problem. Multivariate modeling is considered too advanced, so students never see Simpson's Paradox when working with continuous data.

Although medical tests are an important part of modern life, students are not typically taught about the difference between the quality of a test (on subjects whose disease status is known) and the predictive value of a test (on subjects whose disease status is unknown). As a result, students often commit the Prosecutor's Fallacy (the Base Rate fallacy): concluding that a positive result from a high-quality test is good evidence that the subject is guilty. As a result students cannot properly evaluate the outcomes of tests involving very rare events.

Students are not introduced to the issue of construct validity. Outside psychology, students are seldom introduced to the use of a construct to measure unobservable properties of subjects. IQ is determined by the result of an IQ test; dishonesty is determined by a lie detector test, aptitude is determined by an aptitude test, etc. The validity of these constructs is never raised. Yet evaluating the validity of statistical constructs is central to many of today's debates on social policy. The 'poverty level' and the 'income-gap' are constructs. Statistical constructs must be evaluated.

Students are not exposed to statistical techniques commonly used in making social claims. In *The Bell Curve*, logistic regression is the primary statistical tool. In discussing whether IQ is one thing or several things, factor analysis is the primary statistical tool. Yet logistic regression and factor analysis are seldomly described in an introductory course.

4. Problems in Inferential Statistics

In teaching statistical literacy, inferential statistics overexposes students to the role of chance at the expense of other topics.

Random samples are presented as being statistically representative. As a result, students fail to question whether a particular random sample is representative. Students fail to consider the benefit of stratified random sampling over simple random sampling.

In making statistical inferences, the presence of random sampling is presented as being more important than the kind of study involved. The determination of statistical significance depends on the presence of a random sample; statistical significance is not influenced by whether the random sample was obtained in an experiment or an observational study.

By spending most of the course on probability and inferential statistics, students are lead to believe that chance is more important than bias. Yet "…the most serious threat to the progress of science...comes from bias, not random variation." [Bailar, John. AMSTAT NEWS, Nov., 1997, p. 5]

By overemphasizing statistical inference, students conclude that being statistically significant is sufficient reason for treating an observed association as being one of direct causation. Operationally, statistical significance is treated as if it justified direct causation. If the power of a new drug is statistically significant in a controlled experiment, then the drug is treated as being effective. So if a relationship in an observational study is statistically significant, students mistakenly conclude that this relationship is also one of direct causation. Why else would we spend so much time developing this concept if it did not have this kind of value?

Most introductory texts and courses cover only half the topics needed for statistical literacy. In addition to descriptive statistics and inferential statistics, statistical literacy should include Bayesian statistics and–most of all–evidential statistics.

5. Need for Bayesian Statistics

To be statistically literate when dealing with chance, students need to evaluate the strength of the evidence supporting the explanation of an observed relationship as due to determinate causes as opposed to indeterminate causes (chance). **Bayesian statistics** studies this by treating strength of belief in the truth of a claim as if it were a probability.

In hypothesis testing, Bayesian statistics can be viewed as answering, "How strong is our evidence for saying this outcome is due to chance, given the information we have available?" Suppose we have a statistically significant association with a p-value of 2%. Given our prior belief about the truth of the alternate, what strength of belief is now justified by this new finding? [See Schield, 1996]. Many students believe that if an observed outcome is highly unlikely *if* due to chance, that means the observed outcome is highly unlikely *to be* due to chance.

In confidence intervals, Bayesian statistics can be viewed as answering the question, "How confident could we be in deciding that the unknown population parameter (a fact) is located in this particular confidence interval?" Is betting on a 95% confidence interval the same as betting on drawing a red ball from an urn containing 19 red balls and one green ball? [See Schield, 1997]. In the US, students are seldomly exposed to Bayesian thinking. Yet assessing the strength of an argument is central to critical thinking.

From a classical perspective, these are mistakes:

- the closer a quantile-normal plot of data is to a straight line, the more one is justified in concluding the sample came from a normal distribution.
- the greater the confidence level in a confidence interval, the more one is justified in treating the population parameter as being in that interval.
- the smaller the p-value in a classical hypothesis test, the more one is justified in rejecting the truth of the null hypothesis.
- the p-value is equal to the strength of belief that the null hypothesis is true and the alternate is false. 1-p is the strength of belief that the alternate hypothesis is true and the null is false.
- The greater the relation between two variables the less one is justified in saying it is due to chance.

A Bayesian approach is required to answer the question, "Given our context of knowledge, how likely is it (how strong is the evidence) that this relationship is due to chance?" This is a more difficult question than the inferential question, "How likely is this relationship (or ones even more unlikely) if due to chance?"

6. Need for Evidential Statistics

To be statistically literate, all students need to evaluate the strength of the evidence supporting direct causality between the variables in an observed association. **Evidential statistics** studies the use of statistics as evidence in supporting the truth of non-statistical claims involving causality.

Students often mistakenly conclude that a particular statistical association justifies direct causation.

- Suppose a model of house prices and bathrooms shows an increase of $20,000 for each additional bathroom. Students mistakenly presume if one adds an extra bathroom to a group of random houses, one can expect an increase of $20,000 per house on average.
- Suppose the risk of danger is twice as high in group HIGH as in group LOW. Students mistakenly presume that if random subjects move from HIGH to LOW, they will cut their risk in half.
- Suppose weight and smoking are positively associated. Students mistakenly presume this means if non-smokers gain weight, they have a higher likelihood (chance) of starting smoking.

7. Evidential Statistics and Epistemology

Evidential statistics is a distinct science of method. The sciences of method is a branch of epistemology which studies how we should think in order to avoid error and achieve contextual certainty (Rand, 1965).

Figure 1. A classification of sciences of method

	Kind of argument	
Subject	Deductive-only	Inductive & Deductive
Verbal	Aristotelian Logic Sentential logic	Critical thinking Philosophy of science
Numeric	Math, Probability, Math stats	Evidential statistics

Induction includes causal explanations, generalizations and causal predictions. See Kelley (1994).

"The goal of applied statistics is to help students to form, and think critically about, arguments involving statistics. This construction places statistics further from mathematics and nearer the philosophy of science, critical thinking, practical reasoning and applied epistemology." (Iversen, 1992)

Evidential statistics focuses on the use of practical reasoning (induction) to judge arguments about direct causation. Yet most introductory courses focus entirely on formal reasoning (deduction). So students are ill prepared to evaluate how strongly a statistical premise supports a disputable non-statistical conclusion. Focusing on deduction gives traditional statistics a great benefit ('validity' and 'objectivity') but it does so at a great price ('narrowness' and perhaps even 'irrelevance.')

The relation between evidential statistics and traditional classical statistics is similar to that between macroeconomics and microeconomics. Traditional statistics might be like micro-statistics while evidential statistics might be viewed as macro-statistics. Macro-statistics (evidential statistics) is needed to tie things together–to integrate statistical claims into broader arguments and issues.

8. Evidential Statistics and Kinds of Studies

Evidential statistics focuses on the difference between experiments and observational studies. Students often cannot distinguish one from the other and often mistake an observational study for an experiment.

In traditional statistics, the kind of study is irrelevant as long as the process is random. A two-t test is valid regardless of whether the random data came from an experiment or from an observational study.

In evidential statistics, the kind of study is most relevant – it indicates the amount of support that study can provide in upholding a claim of direct causation.

Inferences about direct causation are much stronger when based on experiments than when based on observational studies. In experiments, researchers can rule out many potentially confounding factors by taking control of their values: directly or statistically. In an observational study, there is always the possibility of an alternate explanation.

Evidential statistics deals primarily with observational studies. It takes much greater skill to evaluate the strength of an argument based on an observational study. It requires wisdom to identify what factors should have been taken into account (controlled for).

Traditional statistics focuses on the strongest studies (e.g., controlled clinical trials) where evidential statistics is least important. Students need experience analyzing statistics obtained from weaker studies that involve less control.

9. Evidential Statistics and Confounding Factors

Evidential statistics deals with potentially confounding factors that can provide alternate explanations to observed associations. The move from association to causation is strengthened by taking into account (controlling for) potentially confounding factors that might better explain the observed association.

In analyzing a statistical association, evidential statistics focuses on the search for alternate explanations that have a greater explanatory power. If the original association vanishes (or reverses) after including other confounding factors, then the comparative statistics do not strongly support the expected conclusion (the original conclusion).

In practice, the strength of any study is strongly influenced by the quality of the study: the extent to which relevant factors are taken into account.

10. Evidential Statistics and Florence Nightingale

Florence Nightingale, the passionate statistician, used evidential statistics to support her claims that improved nursing care would save lives. In 1859, she noted that for every soldier killed in battle in the Crimea, seven died after the battle. But she recognized that this eye-catching statistic gave only weak support for her conclusion. She had no evidence to show that improved nursing care would have made a difference.

Florence Nightingale also presented a more mundane statistic: the death rate of young soldiers in peacetime was double that of the general population. By selection, this study controlled for battle-related deaths and for diseases not prevalent in Great Britain. Although less eye-catching, the mundane (two-fold) statistic actually gave stronger support for her claim than did the eye-catching (seven-fold) statistic. (Brown, p. 44)

Florence Nightingale introduced many techniques designed to take into account (control for) confounding factors. She compared cases (soldiers in barracks) with civilian controls. She noted that mortality statistics should be age-specific and that crude death rates can be misleading. (Johnson and Kotz).

11. Obstacles to Teaching Evidential Statistics

To teach evidential statistics, there are some obstacles that must be overcome.

Need for appreciation of informal induction. Mathematicians and statisticians must appreciate the use of informal induction in trying to identify reasonable conjectures and to create new paradigms.

Need for training. Mathematicians are well trained on deduction–an answer is either right or wrong. Evidential statistics is ultimately inductive. *It will take training for mathematically trained statisticians to teach and test students on evidential statistics.*

Need for domain knowledge. Teaching evidential statistics requires knowledge about specific subjects: their nature, their powers and the interactions with other factors. Evidential statistics cannot be taught as an abstraction–it must be concretized.

Need for course reorientation. In order to add additional topics into a one-semester course, several topics must be omitted. New texts, problems and tests will be required.

12. Benefits of Teaching Evidential Statistics

There are great benefits in teaching evidential statistics in an introductory statistics course.

Uphold importance of statistical literacy. If students are to be able to read and interpret data, they must be statistically literate. Teaching evidential statistics is essential to give them a foundation in critical thinking about statistical claims. The need for evidential statistics is greatest in the humanities. Journalists, policy analysts, historians and philosophers all need to be able to think critically about statistics.

Elevate intellectual status of discipline. By helping students learn to evaluate the strength of inductive arguments, we are helping our students think more efficiently and effectively about explanations and predictions. In so doing, statistics can become an essential part of advanced education.

Reunite Bayesians and Frequentists. *The focus on using statistics in arguments might help frequentists and Bayesians find a more common ground.* Both can focus more on the common goal (strength of belief that a claim is true) and less on their differences (the probability that a fact is true).

Be a guide to public policy. The president of the ASA, David Moore, set the theme of this conference as being "Statistics: A Guide to Public Policy." Evidential statistics can help statistics achieve that role.

13. Conclusion

To be statistically literate, students in all majors need access to an introductory statistics course covers all four areas of statistics: descriptive, inferential, Bayesian and evidential.

To achieve statistical literacy for all, introductory statistics must be expanded to include evidential statistics–the use of statistics as evidence in arguments involving practical reasoning about causality.

Appendix: Related Sources by Topic

Evidence based statistics are discussed in *Say It With Figures (*Zeisel), in *Prove It With Figures: Empirical Methods in Law and Litigation (*Zeisel and Kaye) and in *Statistics, the Principled Argument* (Abelson). Evidence based statistics are commonly used in law. See *Statistics and the Law* by DeGroot, Fienberg and Kadane, *Statistical Reasoning in Law and Policy* by Gastwirth, *The Evolving Role of Statistical Assessments as evidence in the Courts* by Feinberg, and *A Probabilistic Analysis of the Sacco and Vanzetti Evidence* by Kadane and Schum.

For a discussion of the identification problem in the social sciences, see *Identification Problems in the Social Sciences* (Manski) and *Cohort Analysis in Social Research: Beyond the Identification Problem* (Mason and Fienberg, editors). See also *Cross-Level Inference* (Alchen and Shively).

For a discussion of the conceptual primacy of the association between variables, see MacNaughton (1998). The most complex cases occur when statistics in one field (epidemiology, health, or education) are used to support claims in another (law, policy, etc.). See *Phantom Risk, Scientific Inference and the Law* (Foster, Bernstein and Huber).

For examples of the use of critical thinking, see introductory statistics texts by Freedman, Pisani, Purves and Adhikari (1991), Jessica Utts (1991), David Moore (1993), Gary Smith (1998) and Gudmund Iverson (1998).

For a discussion of observational studies and probabilistic causality, see Rosenbaum (1995) and Eerola (1994). For a discussion of epidemiology, see Kelsey, Thompson and Evans (1986) and Bailar (1994). For a discussion of statistics and journalism, see Cohn (1992).

For a discussion of epistemology and logic, see Rand (1965) and Kelley (1994). For a discussion on the philosophy of Science, see Howson and Urbach (1993) and Mayo (1996).

REFERENCES

Abelson, Robert P., (1995). *Statistics as Principled Argument*. Lawerence Earlbaum Assoc. Pub.

Alchen, Christopher H and W. Phillips Shively (1995). *Cross-Level Inference*. University of Chicago Press.

Brown, Pam (1988). *People Who Have Helped the World, Florence Nightingale*. Exley Publications.

Cohn, Victor (1988). *News and Numbers: A Guide to Reporting Statistical Claims and Controversies in Health and Other Fields*. Iowa State University Press

DeGroot, Morris H., Stephen E. Feinberg and Joseph B. Kadane (1994). *Statistics and the Law*, Wiley Interscience 2nd Ed.

Eerola, Mervi (1994). *Probabilistic Causality in Longitudinal Studies.* Springer-Verlag Lecture Notes in Statistics #92.

Fienberg, Stephen E. (1989). *The Evolving Role of Statistical Assessments as Evidence in the Courts* Springer-Verlag

Freedman, David, Robert Pisani, Roger Purves and Ani Adhikari. *Statistics* W.W. Norton & Co., 2nd Ed.

Gastwirth, J.L. (1986). *Statistical Reasoning in Law and Policy*. Academic Press.

Gordon, Florence and Sheldon, Editors. *Statistics for the Twenty-First Century*. The Mathematical Association of America. MAA Notes, Number 26, 1992.

Herrnstein, Richard and Charles Murray (1992). *The Bell Curve*. The Free Press.

Hoaglin, David and Moore, David Editors (1992). *Perspective on Contemporary Statistics.* The Mathematical Association of America. MAA Notes 21.

Howson, Colin and Peter Urbach, (1993). *Scientific Reasoning* 2nd Ed. Open Court Publishing.

Iverson, Gudmund (1984). *Bayesian Statistical Inference*. Sage Publications Series: Quantitative Applications in the Social Sciences.

Iversen, Gudmund (1992). *Two Kinds of Introductory Courses.* Heeding the Call for Change. MAA Notes #29, p. 29.

Johnson, & Kotz (1997). *Leading Personalities in Statistical Sciences.* Wiley-Interscience

Kadane, Joseph B. and David A. Schum (1996). *A Probabilistic Analysis of the Sacco and Vanzetti Evidence*, Springer-Verlag.

Kelley, David (1994). *The Art of Reasoning.* 2nd ed. W.W. Norton & Co.

Kelsey, Jennifer, W. Douglas Thompson, and Alfred Evans (1986). *Methods in Observational Epidemiology*. Oxford University Press.

Macnaughton, Donald (1998). *Eight Features of an Ideal Introductory Statistics Course.* Posted on www.matstat.com

Manski, Charles F. (1995). *Identification Problems in the Social Sciences.* Harvard University Press.

Mason, William M. and Stephen E. Fienberg (1985). *Cohort Analysis in Social Research: Beyond the Identification Problem.* Springer-Verlag.

Mayo, Deborah (1996). *Error and the Growth of Experimental Knowledge.* University of Chicago.

Moore, David S. (1991), *Concepts and Controversies,* 3rd ed. Freeman.

Moore, David & George McCabe (1993). *Introduction to the Practice of Statistics*, 2nd ed., Freeman.

Rand, Ayn (1965). *Introduction to Objectivist Epistemology*. Signet

Rosenbaum, Paul R. (1995). *Observational Studies.* Springer-Verlag.

Schield, Milo (1996). *Using Bayesian Reasoning in Classical Hypothesis Testing.* ASA Proceedings of Section on Statistical Association, p. 274.

Schield, Milo (1997). *Interpreting Statistical Confidence.* ASA Proceedings of Section on Statistical Association, p. 137.

Simon, Julian L. (1996). *The Ultimate Resource 2.* Princeton University Press.

Smith, Gary (1998). *Introduction to Statistical Reasoning.* McGraw Hill

Utts, Jessica (1991), *Seeing Through Statistics*, Duxbury Press.

Zeisel, Hans (1947). *Say It With Figures.* Harper & Row.

Zeisel, Hans and David Kaye (1997). *Prove It With Figures: Empirical Methods in Law and Litigation.* Springer.

Acknowledgments: Linda Schield, Dr. Julie Naylor, Thomas V.V. Burnham and Gerald Kaminski made many helpful suggestions. Drafts of this paper were presented at WesCOTS in Colorado Springs, at Making Statistics More Effective in Schools of Business (MSMESB) in Iowa City, Iowa, and at the Xi'an Statistical Institute in Xi'an China. Dr. Schield can be reached at schield@augsburg.edu. For this paper and related materials, see Dr. Schield's homepage at www.augsburg.edu/ppages/schield.